

Individual Assignment 9

Please download the following files from Canvas:

- Dating App Reviews.csv

The .csv file contains data about a set of 3,623 early reviews for a dating app. Below are the column definitions for this file. Make sure you open up the file and get a sense of the data prior to proceeding.

- Review ID: unique identifier for each review
- Bugs: 1 if the review mentions bugs in the app, 0 if not
- UI/UX: 1 if the review mentions the UI or UX of the app, 0 if not
- Installation: 1 if the review mentions the installation process, 0 if not
- Photos: 1 if the review mentions photos or pictures, 0 if not
- Profile: 1 if the review mentions profiles on the app, 0 if not
- Likes: 1 if the review mentions likes on the app, 0 if not
- Matches: 1 if the review mentions matches on the app, 0 if not
- Chat: 1 if the review mentions chats or conversations on the app, 0 if not
- Spam: 1 if the review mentions spam or fake accounts, 0 if not
- Swiping: 1 if the review mentions swiping behavior, 0 if not
- Meetings: 1 if the review mentions in-person meetings, 0 if not
- Dates: 1 if the review mentions dates, 0 if not
- Friends: 1 if the review mentions friendships, 0 if not
- Fun: 1 if the review mentions fun, 0 if not
- Location: 1 if the review mentions physical locations, 0 if not
- Relationship: 1 if the review mentions romantic relationships, 0 if not

Now open RStudio and create an R Markdown file. Save the .rmd file, ensuring that you save it in a file location that you can easily find again.

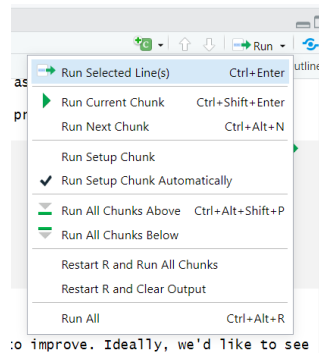
Each of the following items corresponds to a code chunk, or text to record in the R Markdown file. Please make sure to label each item using the numbers provided in this document.

Make sure to run each code chunk after you write it to ensure that there are no errors. Be careful about typos!

If you have not installed the factoextra library yet, do so according to the instructions given in class before beginning this portion of the assignment.

HINT: To ensure consistent results, first complete all of the code chunk portions, and click "Run All". Then use the ensuing output to answer the text answers.

Individual Assignment 9



Item 1 [Code Chunk]. Loading packages: Load the factoextra library.

Item 2 [Code Chunk]. Random seed: Set the random seed to 1234.

Item 3 [Code Chunk]. Importing data: Import the Dating App Reviews.csv file into a dataframe called 'datingappreviews'.

Item 4 [Code Chunk]. Removing unique identifier(s): Remove the unique identifier for each review from the dataframe, and save it into a new dataframe called 'datingappreviews_subset'.

Item 5a [Code Chunk]. Choosing optimal k: Using datingappreviews_subset, plot the within-cluster sum of squares for various numbers of clusters. Make sure the y-axis goes from 0 to 11,500.

Item 5b [Text]. Choosing optimal k: Using the elbow method and the graph you created for Item 5a, identify what the optimal number of clusters is. Justify your answer.

Item 6 [Code Chunk]. K-means clustering: Run k-means clustering on datingappreviews_subset using the number of clusters you identified in Item 5b.

Item 7a [Code Chunk]. Cluster centroids: Output the resulting cluster centroids.

Item 7b [Text]. Cluster centroids: Using the cluster centroids, summarize the types of reviews that ended up in Cluster 5.

Item 8a [Code Chunk]. Within-cluster sum of squares: Calculate the within cluster sum of squares.

Item 8b [Text]. Within-cluster sum of squares: Interpret the within-cluster sum of squares.

Item 9a [Code Chunk]. Between-cluster sum of squares: Calculate the between cluster sum of squares.

Item 9b [Text]. Between-cluster sum of squares: Interpret the between-cluster sum of squares.

Item 10a [Code Chunk]. Explained variance: Calculate the explained variance of the clustering.

Item 10b [Text]. Explained variance: Interpret the explained variance metric.

Save your file, then knit and export your R Markdown file as an HTML file. Upload the HTML file to Canvas to complete the assignment.

If you are unable to knit your .rmd file due to errors, make sure that you go back and test your code chunks individually. If you are ultimately unable to figure out how to solve these errors, save the .rmd file and upload that instead of the HTML file for partial credit.

Individual Assignment 9

If you are unable to find your exported HTML file, consult Canvas for instructions.